

Pierre-André BUVET  
Centre Tesnière  
Université de Franche-Comté

Michel MATHIEU-COLAS  
LLI, CNRS – INaLF (UMR 7546)  
Université Paris-Nord

## **LES CHAMPS *DOMAINE* ET *SOUS-DOMAINE* DANS LES DICTIONNAIRES ÉLECTRONIQUES\***

Les dictionnaires électroniques élaborés au LLI<sup>1</sup> sont destinés au traitement automatique de la langue ; leur structuration correspond à celle des bases de données, avec des champs préalablement et clairement définis. Chaque entrée du dictionnaire décrit un emploi de mot auquel sont attachées un certain nombre d'informations standardisées :

— de nature morphologique (structure du mot, flexion, variantes graphiques) ;

— de nature syntactico-sémantique ; ainsi, les substantifs sont décrits, dans le champ T, en termes de traits<sup>2</sup> et, dans le champ C, par des descripteurs plus puissants, les classes d'objets<sup>3</sup>.

D'autres informations du second type peuvent être également spécifiées ; entre autres, les indications de domaine (champ **D**), de sous-domaine (champ **SD**) et de registre (champ **R**). Le rattachement variable d'un mot à un domaine en fonction du degré de précision que l'on attribue à ce type d'information explique la distinction entre les champs **D** et **SD** ; par exemple, le composé *droite sécante* peut être considéré, selon les lexicographes, comme relevant des mathématiques ou, plus spécifiquement, de la géométrie. Tout dépend, en l'occurrence, du point de vue adopté et des applications envisagées. Dans les dictionnaires du LLI, nous choisissons de coder les deux types d'informations respectivement dans les champs **D** et **SD**. Leur articulation, on le verra, n'est pas toujours

---

\* Nous présentons ici la synthèse des travaux d'un groupe de réflexion (Michel MATHIEU-COLAS, Jean-Claude ANDRÉ, Pierre-André BUVET et Chantal GIRARDIN avec la participation de Jacques BOISSY et Danielle CANDEL) sur le codage des domaines dans les dictionnaires électroniques, Villetaneuse 1996-1997.

<sup>1</sup> Laboratoire de Linguistique Informatique, dirigé par Gaston Gross (UMR 7546, CNRS-INaLF, Université Paris13, Villetaneuse)

<sup>2</sup> Les traits utilisés permettent de caractériser des noms relatifs à des humains, des animaux, des végétaux, des inanimés concrets, des locatifs, des temps, des événements, des actions, des états.

<sup>3</sup> Sur la notion de classe d'objets, cf. G. Gross 1992, 1994, 1995, D. Le Pesant 1994, D. Le Pesant et M. Mathieu-Colas 1998, M. Mathieu-Colas 1994, 1998.

aussi simple que dans l'exemple précédent (cf. 2.2.). Par ailleurs, la relation entre les indications de domaine et celles de registre diffère de la pratique lexicographique traditionnelle en établissant une distinction explicite entre marque d'emploi et marque d'usage (cf. 1.2. et 2.1.2.).

Avant de détailler les principes du codage des indications de domaines et sous-domaines que nous proposons, nous examinerons la notion même de domaine, tant dans la tradition lexicographique et terminologique que dans les dictionnaires électroniques du LLI ; la spécificité de la nomenclature que nous présentons sera également traitée. Nous discuterons aussi de l'articulation entre les classes d'objets et les domaines, ainsi que de l'intérêt des indications de domaine pour l'élaboration de lexiques spécialisés et pour le traitement automatique de l'ambiguïté.

## **1. La notion de domaine**

### **1.1. La tradition lexicographique et la pratique terminologique**

Une pratique lexicographique des plus courantes est de spécifier à quelle langue spécialisée appartient une unité lexicale<sup>4</sup> ; il en résulte, comme dans le cas du *Robert méthodique*, « que l'on trouve avant la définition de très nombreux mots, sens ou expressions une marque de niveau de langue qui précise la valeur de l'emploi [...] dans un domaine d'activité (*techn.* : mot du langage technique ; *admin.* : mot employé dans la langue écrite de l'administration ; *didact.* : terme de la langue savante, que l'on peut trouver dans un traité, dans un cours, mais que l'on n'utilise pas dans le langage courant) » J. Rey-Debove et C. Bellefonds 1989 : X-XI.

Par ailleurs, la notion de domaine est constitutive de la terminologie<sup>5</sup>, définie comme « un système de termes employés pour exprimer les notions propres à une science donnée » J. Marouzeau 1969, « un ensemble de termes [...] par lesquels toute science désigne les notions qui lui sont utiles » J. Dubois & alii 1973, « l'ensemble des termes qui renvoient aux concepts et aux objets afférents à un domaine de connaissance ou d'activité humaine » R. Galisson et D. Coste 1976 ou « un ensemble d'expressions

---

<sup>4</sup> Les dictionnaires anciens n'ignorent pas cette pratique ; ainsi, dans le premier *Dictionnaire de l'Académie* (1694), un des emplois du nom *expression* est décrit « en termes de Peinture et de Sculpture » ; cf. C. Girardin 1994.

<sup>5</sup> Le mot *terminologie* est ambigu puisqu'il désigne à la fois une discipline et l'objet de cette discipline. En tant que discipline la terminologie est définie comme "l'étude scientifique des notions et des termes en usage dans les langues de spécialités" ISO-1807 1990. Les définitions ci-dessus concernent l'objet de la discipline.

dénommant dans une langue naturelle des notions relevant d'un domaine de connaissances fortement thématized » P. Lerat 1995. La description d'un terme<sup>6</sup> apparaît donc indissociable de l'indication de son domaine de rattachement.

Les définitions des langues scientifiques et techniques que proposent la plupart des terminologues sont assez bien représentées par les deux propositions suivantes :

« La langue de spécialité est, dans ses textes, le trésor du savoir disponible et linguistiquement analysable et, dans ses ressources, l'instrument du savoir futur » R. Kocourek 1991 : 42.

« On peut définir [une langue spécialisée<sup>7</sup>] comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées » P. Lerat 1995 : 21.

La conception terminologique des langues scientifiques et techniques relève donc, en premier lieu, de l'usage qu'en font les spécialistes. Dans la pratique lexicographique, l'application de ce point de vue consiste souvent à coder des indications de domaine uniquement pour les unités lexicales qui sont stricto sensu des termes scientifiques ou techniques. Ainsi, dans le *Grand Dictionnaire encyclopédique Larousse* 1985, seul le substantif *bout* reçoit une marque (marine) ; par contre, cette indication est exclue pour son synonyme *cordage* dans la mesure où la seconde dénomination ne figure pas dans des nomenclatures spécialisées. La réalité lexicographique est souvent plus confuse comme l'atteste le traitement disparate des substantifs *jaunisse* et *ictère* dans différents dictionnaires<sup>8</sup> (cf. M. Glatigny 1994) :

	<i>jaunisse</i>	<i>ictère</i>
<i>NPR</i>	COUR	MÉD.
<i>Lexis</i>	Méd.	Pathol.
<i>RC</i>	Méd	
<i>PLI</i> 1996		MÉD.
<i>TLF</i>		PATHOL

<sup>6</sup> Par terme, on comprendra ici "dénomination spécialisée" P. Lerat 1995 plutôt qu'une "unité définie dans les textes de spécialité" R. Kocourek 1991.

<sup>7</sup> Selon P. Lerat 1995, on doit parler de langue spécialisée plutôt que de langue de spécialité, la langue spécialisée étant définie comme "une langue naturelle considérée en tant que vecteur de connaissances spécialisées".

<sup>8</sup> *NPR* : *Nouveau Petit Robert*, *RC* : *Robert-Collins*, *PLI* : *Petit Larousse illustré*, *TLF* : *Trésor de la Langue Française*.

## 1.2. Elargissement de la notion

Une autre option est de considérer que « les indications de domaine n'indiquent pas seulement le champ d'expérience dont relève le mot (cf. *agentif*, **ling.**) mais aussi à propos duquel on l'utilise » B. Quemada & alii 1984. C'est ce second point de vue qui prévaut dans les dictionnaires électroniques du LLI ; ainsi, dans le champ **D**, les substantifs *bout* et *cordage* sont l'un et l'autre présentés comme relevant du domaine nautique tandis que les noms *ictère* et *jaunisse* sont l'un et l'autre considérés comme appartenant au domaine de la médecine. Le fait qu'une unité lexicale appartienne à une nomenclature scientifique ou technique (comme dans le cas de *bout* et *ictère*) n'est pas pour autant négligé ; l'information relative aux langues spécialisées est enregistrée dans le champ **R** :

	<b>D</b>	<b>R</b>
<i>bout</i>	<b>élevage</b>	langue spécialisée
<i>cordage</i>	<b>élevage</b>	langue générale
<i>ictère</i>	<b>médecine</b>	langue spécialisée
<i>jaunisse</i>	<b>médecine</b>	langue générale

Par rapport à la pratique lexicographique courante, c'est donc un double élargissement que nous proposons dans la mesure où nous précisons, dans le champ **D**, le secteur d'activité auquel renvoie le mot décrit et, dans le champ **R**, l'éventuel usage spécialisé qui en est fait. Le choix de décrire les unités lexicales à la fois du point de vue de leur référent<sup>9</sup> (en termes de domaines) et du point de vue de leur usage (langue spécialisée vs langue générale) permet un codage plus précis<sup>10</sup>.

Les dictionnaires du LLI se différencient également par les principes d'élaboration de la nomenclature. Il ne s'agit pas pour nous d'établir *a priori* un système conceptuel hiérarchisé, comme celui qu'utilisent les bibliothèques pour le catalogage (classifications Dewey, CDU, etc.). Les indications de domaine ont à nos yeux pour fonction essentielle de mieux décrire le lexique (parmi d'autres descripteurs) et non de rendre compte de réalités extra-linguistiques. Ce choix a deux conséquences.

---

<sup>9</sup> Cette pratique est peu courante à l'exception des dictionnaires bilingues : le *Robert-Collins* code « jur » et « comm » pour deux emplois du nom *acheteur* et « mil » et « scoutisme » pour deux emplois du nom *troupe*.

<sup>10</sup> Notons qu'il y a souvent coïncidence des deux approches. Le composé *cap-hornier* relève du langage des marins, en même temps qu'il désigne une réalité liée à la mer.

En premier lieu, la liste des domaines ou sous-domaines du LLI ne s'apparente pas à une nomenclature hiérarchisée. L'absence de hiérarchie facilite le codage dans la mesure où une nomenclature conceptuelle, par nature arborescente, interdit de traiter des cas de double rattachement comme, par exemple, les unités *soutane*, *chanteur de charme* et *chapeau du capitaine* (pourcentage sur le fret perçu par le capitaine) qui, respectivement, relèvent des deux domaines de l'habillement et de la religion, de la musique et du spectacle, du droit et de la marine. Une pratique plus empirique permet de décrire ce type de situations, en rendant compte de l'appartenance simultanée d'une unité lexicale donnée à différents domaines<sup>11</sup> :

unité lexicale	domaine
<i>soutane</i>	<b>habillement ; religion</b>
<i>chanteur de charme</i>	<b>musique ; spectacle</b>
<i>chapeau du capitaine</i>	<b>droit et justice ; nautique</b>

Par ailleurs, la description systématique des domaines a nécessité le développement de nouvelles rubriques. Aucun dictionnaire usuel n'attribue de domaine spécifique à des mots comme *assurance-décès*, *tente-caravane*, *birth-control*, *stylo-bille*, *croque-mort*, *living-room*, *colis piégé*, etc. Cela tient en partie à l'insuffisance des nomenclatures. La création de nouveaux sous-domaines permet de prendre en charge un plus grand nombre d'unités :

unité lexicale	sous-domaine
<i>assurance-décès</i>	<b>assurances</b>
<i>tente-caravane</i>	<b>camping</b>
<i>birth-control</i>	<b>contraception</b>
<i>stylo-bille</i>	<b>écriture</b>
<i>croque-mort</i>	<b>funérailles</b>
<i>living-room</i>	<b>habitat</b>
<i>colis piégé</i>	<b>terrorisme</b>

---

<sup>11</sup> De la même façon qu'une unité lexicale peut relever de plusieurs classes d'objets ; ainsi *pétrolier* appartient à fois à la classe <**contenant**> et à la classe <**moyen de transport**>.

Des extraits des listes des domaines et des sous-domaines utilisés dans les dictionnaires du LLI, ainsi que leurs codes, figurent en annexe.

### 1.3. Domaines et classes d'objets

Si la description des unités lexicales, dans les dictionnaires du LLI, s'appuie principalement sur les classes d'objets, les indications de domaine n'en demeurent pas moins indispensables : les noms *banc d'étirage*, *bétonneuse* et *bineuse* ont en commun d'appartenir à la classe d'objets <**machine**> et diffèrent seulement par leur appartenance respective aux domaines du textile, de la construction et de l'agriculture. De même, la plupart des emplois de *pied-de-biche* relèvent de la classe <**outil**> et se distinguent par leur rattachement aux domaines de l'armement, de l'horlogerie, de la manutention, etc.

Il est vrai que les informations spécifiées en termes de domaines sont parfois redondantes par rapport aux classes d'objets : les noms propres *Nationale 7*, *Francilienne*, *autoroute de l'Ouest* de la classe d'objets <**voie**> relèvent tous du domaine du transport. Cette redondance n'est pourtant pas une règle générale : certains des emplois des noms *glacage*, *nappage* et *pochage* sont relatifs au domaine de l'alimentation sans appartenir, pour autant, à la classe <**aliment**> ; de même, *dorsi-ventralité* ou *pseudo-fécondation* relèvent de la botanique sans désigner des plantes... Du point de vue de l'extension, les classes d'objets et les domaines ne coïncident pas, mais sont en rapport d'intersection.

### 1.4. Domaines et extraction automatique

Le codage des domaines dans un dictionnaire électronique permet d'identifier automatiquement le vocabulaire relatif à thème donné : il suffit d'extraire toutes les unités lexicales associées à un domaine spécifique. Par exemple, une requête portant sur les termes de marine dans le dictionnaire électronique des composés *N de N* (en cours d'élaboration) donne instantanément la sélection suivante :

*aire du vent, aumônier de la marine, avarie de mer, acte de francisation, amarre de bout, amure de foc, amure de trinquette, amure de voile, ancre d'empenelle, ancre de bossoir ...*

Il devient ainsi possible d'envisager, à partir de l'ensemble des données enregistrées dans les dictionnaires, la constitution automatique de lexiques spécialisés ou, plus généralement, relatifs à un domaine d'activité particulier (du camping aux arts martiaux...).

## 1.5. Domaines et levée d'ambiguïté

Les indications de domaines d'emploi figurant dans les dictionnaires électroniques peuvent aussi concourir, dans le cadre d'analyses automatiques de textes, à désambiguïser des unités lexicales. La désambiguïstation de constructions nominales *N de N* polysémiques en étiquetant, à l'aide d'un dictionnaire électronique, les domaines des substantifs de textes numérisés a été expérimentée (cf. P.-A. Buvet 1996) ; l'hypothèse de travail était que l'interprétation adéquate devait résulter de la corrélation entre les emplois spécialisés des substantifs environnants et celui de la séquence a priori ambiguë. Les résultats obtenus font état de la possibilité, pour un système automatique, de lever certaines ambiguïtés lexicales à partir des informations de domaines telles qu'elles figurent dans les dictionnaires électroniques. Cette étude rend compte également de situations où les seules informations de domaines rattachées aux substantifs sont insuffisantes pour réduire automatiquement la polysémie des structures *N de N*. Des dispositifs effectuant des analyses morpho-syntaxiques et distributionnelles fines sont donc indispensables. On peut néanmoins considérer qu'un système performant destiné à lever automatiquement des ambiguïtés dans des textes peut adjoindre à ses outils d'analyse un procédé d'étiquetage des domaines d'emploi des substantifs par le biais d'un dictionnaire électronique prenant en compte les domaines.

Notons qu'il est également possible d'envisager pour le traitement automatique des textes spécialisés une déclaration préalable de domaine qui a pour effet de sélectionner certains emplois de mots<sup>12</sup>.

## 2. Principes de codage

Dans un premier temps, nous présenterons les notions de « transdomaine », de « spécificateur temporel » et de « spécificateur géographique » dont la caractéristique commune est une économie de représentation pour le marquage des domaines. Nous discuterons également des procédures particulières du codage des différents jeux et sports (spécificateurs thématiques). Dans un deuxième temps, nous traiterons des articulations entre le champ **D** et le champ **SD**.

---

<sup>12</sup> Ce type de procédure est déjà opérationnel dans le système de traduction automatique SYSTRAN. La déclaration de domaine <topical glossary> donne priorité à l'une des traductions en cas de polysémie.

## 2.1. Les compléments des sous-domaines

Cette section est consacrée à des informations plus ou moins standardisées qui ont la particularité de compléter différentes spécifications de domaines ou sous-domaines.

### 2.1.1. Les transdomaines<sup>13</sup>

Pour rendre compte de l'intérêt de la notion de transdomaine, nous prendrons l'exemple de la marque **histoire**. Elle figure naturellement dans la liste des domaines et recouvre, à ce titre, tout le vocabulaire des sous-domaines suivants : **Antiquité, archéologie, chevalerie, chronologie, diplomatique, droit coutumier, droit féodal, droit romain, féodalité, histoire coloniale, histoire contemporaine, numismatique, paléographie, préhistoire, protohistoire, sigillographie**, ainsi que le vocabulaire employé en histoire d'une façon très générale ou au contraire très spécifique (cf. *infra*).

Mais il existe aussi des mots propres à l'histoire d'un domaine particulier : par exemple, histoire de l'art, histoire du cinéma, histoire du costume, histoire ecclésiastique, histoire littéraire, histoire de la médecine, histoire militaire, histoire du mobilier, histoire des mœurs, histoire de la musique, histoire de la peinture, histoire des sciences, histoire des techniques, histoire du théâtre, histoire des transports. Cette liste n'est pas exhaustive dans la mesure où toute activité humaine est susceptible de générer un vocabulaire prenant en compte la dimension historique de cette activité. Aussi, plutôt que d'intégrer toutes les rubriques dans la liste des sous-domaines, nous avons pris le parti de considérer **histoire** (codé :HIST) comme un « transdomaine », c'est-à-dire un élément permettant de spécifier, pour une unité lexicale donnée, un aspect particulier de l'information apportée par l'indication de domaine (en l'occurrence sa dimension historique) :

unité lexicale	domaine	(sous-)domaine+transdomaine
<i>ciné-œil</i> <i>cinéma muet</i>	<b>cinéma</b>	<b>cinéma:HIST</b>
<i>sous-avoué</i> <i>bréviaire de Quignonez</i>	<b>religions</b>	<b>catholicisme:HIST</b>

---

<sup>13</sup> Les transdomaines sont représentés par quatre lettres majuscules précédées de deux points.



unité lexicale	domaine	(sous-)domaine+transdomaine
<i>voix-de-ville</i> <i>symbolisme</i>	<b>littérature</b>	<b>littérature:HIST</b>
<i>ramasse-piétons</i> <i>voiture de troisième classe</i>	<b>transports</b>	<b>transports:HIST</b>

L'ensemble des transdomaines peut s'appliquer à toutes sortes de domaines (ou sous-domaines), l'association des deux marques constituant par elle-même un sous-ensemble du domaine (ou sous-domaine) concerné ; ainsi, le transdomaine **industrie** (codé **:INDU**), appliqué à des marques comme **alimentation**, **textile**, **caoutchouc** suffit à spécifier le vocabulaire de « l'industrie alimentaire », de « l'industrie textile » ou de « l'industrie du caoutchouc »<sup>14</sup> :

unité lexicale	domaine	(sous-)domaine + transdomaine
<i>radio-appertisation</i>	<b>alimentation</b>	<b>alimentation:INDU</b>
<i>texturer</i>	<b>textile</b>	<b>textile:INDU</b>
<i>caoutchoutier</i>	<b>matériaux</b>	<b>caoutchouc:INDU</b>

Certains transdomaines sont plus originaux ; c'est le cas d'indications comme **théorie** (code **:THEO**) ou **application** (code **:APPL**), qui permettent de représenter de façon économique des rubriques telles que « informatique théorique » (**informatique:THEO**) ou « linguistique appliquée » (**linguistique:APPL**). De même le code **:EQUI**, désignant l'**équipement** (et plus généralement toute forme de matériel : accessoires, instruments, machines, mobilier, etc.) est applicable à de nombreux domaines et sous-domaines, du matériel d'écriture (**écriture:EQUI**) aux instruments agricoles (**agriculture:EQUI**) en passant par les articles de fumeur (**tabac:EQUI**) ... Par ailleurs, les transdomaines **général** (codé **:GENE**) ou **divers** (codé **:DIVE**) se distinguent de tous les autres dans la mesure où, associés à un domaine (et non à un sous-domaine), ils permettent d'indiquer, dans le cas de **général**, s'il s'agit d'un vocabulaire, relatif au domaine, d'une portée très générale ou, dans le cas de **divers**, d'unités lexicales dont l'emploi est très spécifique sans relever pour autant d'un sous-domaine répertorié :

---

<sup>14</sup> On notera que les expressions (sous-)domaine + transdomaine équivalent le plus souvent à des composés du type *NAdj* ou *N de N* (**littérature:HIST** = *histoire littéraire* ou *histoire de la littérature*).

unité lexicale	domaine	sous-domaine
<i>théorème</i>	<b>mathématiques</b>	<b>mathématiques:GENE</b>
<i>démarche empirique</i>	<b>sciences</b>	<b>sciences:GENE</b>
<i>homme-orchestre</i>	<b>musique</b>	<b>musique:DIVE</b>
<i>voie d'eau</i>	<b>nautique</b>	<b>nautique:DIVE</b>

Notons, comme l'indiquent les exemples ci-dessus, que les transdomaines figurent uniquement dans le champ **SD**.

### 2.1.2. Les spécificateurs géographiques et les spécificateurs temporels

Pour certaines unités lexicales, des informations d'ordre géographique et temporel relatives aux domaines ou aux sous-domaines peuvent être précisées par le biais de spécificateurs appropriés. A l'instar de l'utilisation des transdomaines, celle des spécificateurs géographiques et des spécificateurs temporels permet une économie de représentation pour le marquage des domaines ou des sous-domaines ; ainsi, pour caractériser le vocabulaire relatif à l'architecture antique, l'habillement ancien, la métrique française, la mythologie grecque ou la philosophie scolastique, plutôt que d'augmenter la liste des sous-domaines, on procède de la façon suivante :

unité lexicale	domaine	(sous-)domaine + spécificateur géographique + spécificateur temporel
<i>péristyle</i>	<b>architecture</b>	<b>architecture (Antiquité)</b>
<i>braie</i>	<b>habillement</b>	<b>habillement (anc.)</b>
<i>alexandrin</i>	<b>littérature</b>	<b>métrique (France)</b>
<i>centaure</i>	<b>croyances et idéologies</b>	<b>mythologie (Grèce, Antiquité)</b>

Ces spécificateurs ne dépendent pas d'une nomenclature préétablie mais sont librement introduits en fonction des besoins de la description. Seules exceptions : les codes **anc.** et **contemp.**, qui permettent de situer les unités dans le temps en l'absence d'autres précisions.

Rappelons que, dans les champs **D** et **SD**, on indique le secteur d'activité auquel renvoie le mot décrit et non les marques d'usage qui caractérisent les locuteurs (celles-ci sont indiquées dans le champ **R**). En d'autres termes, il ne faut pas confondre les spécificateurs géographiques avec des marques lexicographiques comme **régionalisme**, **belgicisme**,

**canadianisme** ou les spécificateurs temporels avec des indications comme **vieux, néologisme**.

Comparons, de ce point vue, *football américain* et *ballon-panier*. L'un désigne une variété de sport spécifiquement étrangère (spécificateur géographique), tandis que l'autre n'est spécifique que sur le plan des locuteurs (canadianisme pour « basket-ball ») :

unité lexicale	sous-domaine	registre
<i>football américain</i>	<b>sports (USA)</b>	
<i>ballon-panier</i>	<b>basket-ball</b>	<b>Canada</b>

De même, sur le plan temporel, des mots comme *troubadour*, *Ponant* (au sens de « océan Atlantique ») et *bateau-canon* ne reçoivent pas le même traitement. Le premier se réfère à une époque ancienne (spécificateur temporel), au contraire du deuxième, dont seul l'usage est archaïque, cependant que le troisième accuse un double vieillissement, tant du point de vue de l'usage que du point de vue du référent<sup>15</sup> :

unité lexicale	sous-domaine	registre
<i>troubadour</i>	<b>poésie:HIST (Moyen Age)</b>	
<i>Ponant</i>	<b>géographie</b>	<b>vx</b>
<i>bateau-canon</i>	<b>marine (anc.)</b>	<b>vx</b>

### 2.1.3. Modalités de codage

L'association d'un transdomaine à un domaine ou à un sous-domaine est symbolisée par le signe des deux-points. Par contre, ce sont des parenthèses ouvrantes et fermantes qui indiquent qu'un spécificateur géographique ou un spécificateur temporel complète un domaine ou un sous-domaine. Les spécificateurs géographiques et les spécificateurs temporels peuvent se combiner entre eux, d'une part, avec les transdomaines, d'autre part. Dans tous les cas de figure, ces différentes relations sont spécifiées uniquement dans le champ **SD**. Le tableau ci-dessous récapitule différentes possibilités d'utilisation des éventuels compléments des domaines ou des sous-domaines :

<sup>15</sup> Voir la définition de *bateau-canon* dans le GDEL : "*Nom donné au XIX<sup>e</sup> siècle à quelques navires légers expérimentaux, dotés d'une pièce de gros calibre située à l'avant. (Ils ont été rapidement abandonnés en raison de leur instabilité et de leur manque de rayon d'action.)*" ; nous soulignons.

unité lexicale	domaine	(sous-)domaine + transdomaine + spécificateur géographique + spécificateur temporel
<i>Ars nova</i>	<b>musique</b>	<b>musique:HIST</b>
<i>chef d'orchestre</i>	<b>musique</b>	<b>musique:GENE</b>
<i>chorus</i>	<b>musique</b>	<b>jazz:TECH</b>
<i>conservatoire</i>	<b>musique</b>	<b>musique:ENSE</b>
<i>contrepoint</i>	<b>musique</b>	<b>musique:TECH</b>
<i>double-croche</i>	<b>musique</b>	<b>notation musicale</b>
<i>free jazz</i>	<b>musique</b>	<b>jazz</b>
<i>galoubet</i>	<b>musique</b>	<b>musique (Provence)</b>
<i>gamelan</i>	<b>musique</b>	<b>musique (Indonésie)</b>
<i>gong</i>	<b>musique</b>	<b>organologie</b>
<i>gymel</i>	<b>musique</b>	<b>musique:HIST (Angleterre, 13e s.)</b>
<i>organum</i>	<b>musique</b>	<b>musique:HIST (Moyen Age)</b>
<i>semi-brève</i>	<b>musique</b>	<b>notation musicale:HIST (Moyen Age)</b>

#### 2.1.4. Les spécificateurs thématiques

Certains domaines comportent un grand nombre de spécialisations peu productives lexicalement. Plutôt que de les ajouter à la liste des sous-domaines, au risque de l'agrandir d'une façon disproportionnée, nous choisissons de les indiquer entre crochets. C'est le cas, notamment, pour un grand nombre de jeux et de sports :

unité lexicale	domaine	sous-domaine
<i>contre-jan</i>	<b>jeux</b>	<b>jeux [trictrac]</b>
<i>demi-casei</i>	<b>jeux</b>	<b>jeux [jacquet]</b>
<i>dix de der</i>	<b>jeux</b>	<b>cartes [belote]</b>
<i>joko-garbi</i>	<b>sports</b>	<b>sports [pelote basque]</b>
<i>brise-reins</i>	<b>sports</b>	<b>sports [lutte]</b>
<i>épaulé-jeté</i>	<b>sports</b>	<b>sports [haltérophilie]</b>
<i>extenseur-exerciseur</i>	<b>sports</b>	<b>sports [culture physique]</b>

Seuls échappent à ce codage les jeux (tels les échecs) ou les sports (par exemple le cyclisme) qui se distinguent par l'importance quantitative de leur lexique.

Le même système peut s'appliquer à d'autres domaines, de la religion à la mécanique en passant par l'éducation :

unité lexicale	domaine	sous-domaine
<i>digambara</i>	<b>religions</b>	<b>religions [jainisme]</b>
<i>méthode Freire</i>	<b>éducation</b>	<b>formation [alphabétisation]</b>
<i>presse à balancier</i>	<b>mécanique</b>	<b>mécanique:INDU [emboutissage]</b>

L'intérêt de ces spécificateurs libres est de ne pas perdre d'information. Il n'est pas exclu que certains d'entre eux puissent être intégrés ultérieurement à la liste des sous-domaines.

## 2.2. Articulation entre le champ D et le champ SD

Nous avons vu comment les indications de domaine (champs **D** et **SD**) s'articulent, d'une part, avec les niveaux de langue (champ **R**), d'autre part, avec les classes d'objets (champ **C**). Il sera question ici de l'articulation entre les domaines et les sous-domaines.

Nous rejetons, rappelons-le, toute hiérarchie *préétablie* du type :

domaine A = sous-domaines A1, A2, A3, A4...  
 domaine B = sous-domaines B1, B2, B3, B4...  
 domaine C = sous-domaines C1, C2, C3, C4...

Satisfaisant du point de vue logique, un tel système s'adapte mal à la réalité lexicale ou, plus exactement, ne s'y adapte que partiellement. Si de nombreux sous-domaines se laissent rattacher de manière univoque à un domaine englobant (**algèbre** sous **mathématiques**, **ornithologie** sous **zoologie**, **catholicisme** sous **religions**, etc.), il en est d'autres qui échappent à ce type d'arborescence. Il peut y avoir des rattachements multiples (le **droit commercial** relève simultanément du **droit** et du **commerce**), comme des absences de rattachement (sous quel domaine placer, sans artifice, les rubriques **harnachement** ou **généalogie** ?) A cela s'ajoute le fait qu'un même sous-domaine peut mettre à contribution, selon les unités à coder, des domaines différents (les termes relatifs à l'**hygiène** seront à mettre, selon les cas, en rapport avec la **médecine**, l'**éducation**, l'**environnement**...). D'autres mots, aussi bien, peuvent chevaucher plusieurs sous-domaines (la photo-interprétation ressortit à la fois à la **photographie aérienne** et à la **cartographie**).

Compte tenu de cette complexité, il nous semble nécessaire d'admettre, entre les champs **D** et **SD**, une articulation très souple, ou pour mieux dire une relative autonomie. Les exemples qui suivent illustrent les principaux types de relations que nous avons pu observer.

1) Un sous-domaine et un domaine :

entrée	domaine	sous-domaine
<i>demi-treillis</i>	<b>mathématiques</b>	<b>algèbre</b>
<i>général d'armée</i>	<b>militaire</b>	<b>armée de terre</b>
<i>rétinite</i>	<b>médecine</b>	<b>ophtalmologie</b>

La relation entre les deux instances n'est pas figée, mais peut dépendre des unités lexicales : *fusil de chasse* et *fusil mitrailleur* (SD:**armes**) se rattachent respectivement à la **chasse** et au domaine **militaire**. De même les **baux et locations** peuvent relever, selon les cas, du **commerce** ou de l'**habitation**.

entrée	domaine	sous-domaine
<i>fusil de chasse</i>	<b>chasse</b>	<b>Armes</b>
<i>fusil-mitrailleur</i>	<b>militaire</b>	<b>Armes</b>
<i>bail commercial</i>	<b>commerce</b>	<b>baux et locations</b>
<i>bail d'habitation</i>	<b>habitat</b>	<b>baux et locations</b>

2) Un sous-domaine et plusieurs domaines simultanés :

entrée	domaine	sous-domaine
<i>demi-cellule</i>	<b>chimie ; électricité</b>	<b>électrochimie</b>
<i>drogue dure</i>	<b>médecine ; société</b>	<b>toxicomanie</b>
<i>hors-cote</i>	<b>économie ; finance</b>	<b>Bourse</b>
<i>tribunal de commerce</i>	<b>commerce ; droit et justice</b>	<b>droit commercial</b>

Notons qu'un sous-domaine déterminé peut se rattacher de manière optionnelle à des domaines complémentaires selon les unités lexicales : si *examen psychiatrique* et *expertise psychiatrique* (sous-domaine : **psychiatrie**) relèvent tous deux de la **médecine**, le deuxième terme, et lui seul, se rattache à l'ensemble **droit et justice**<sup>16</sup> :

<sup>16</sup> Dans la table de correspondance entre domaines et sous-domaines, nous indiquons les principales options entre parenthèses :

entrée	domaine	sous-domaine
<i>examen psychiatrique</i>	<b>médecine</b>	<b>psychiatrie</b>
<i>expertise psychiatrique</i>	<b>médecine ; droit et justice</b>	<b>psychiatrie</b>

3) Certains sous-domaines ne se rattachent à aucun domaine clairement identifiable ; dans de tels cas, très peu nombreux (la liste est indiquée ci-dessous), on parlera de domaine zéro (0) :

domaine	sous-domaine
0	<b>accidents</b>
0	<b>bureau (articles de)</b>
0	<b>Cartonnage</b>
0	<b>Cryptographie</b>
0	<b>dessin industriel</b>
0	<b>Généalogie</b>
0	<b>gommes et résines</b>
0	<b>Harnachement</b>
0	<b>Lunetterie</b>
0	<b>Nettoyage</b>
0	<b>Papier</b>
0	<b>Physionomie</b>
0	<b>Salines</b>
0	<b>tabac</b>
0	<b>yoga</b>

4) A l'inverse, pour certains mots, il est difficile d'identifier un sous-domaine précis. Nous reproduisons le cas échéant la marque de domaine, assortie d'un des deux transdomaines **:GENE** ou **:DIVE** (cf. *supra* 2.1.1.) :

---

SOUS-DOMAINE	DOMAINE
<b>anthropométrie</b>	<b>anthropologie + biologie (+ droit)</b>
<b>broderie</b>	<b>textile (+ travaux et équipements ménagers + habillement + ...)</b>
<b>hygiène</b>	<b>médecine (+ éducation + environnement et urbanisme + ...)</b>

entrée	domaine	sous-domaine
<i>parc industriel</i>	<b>industrie</b>	<b>industrie:GENE</b>
<i>chasse à l'affût</i>	<b>chasse</b>	<b>chasse:DIVE</b>

5) Un même unité peut relever simultanément, et pour un même emploi, de plusieurs sous-domaines (à distinguer, naturellement, des cas de polysémie, pour lesquels nous choisissons de dégrouper les entrées). On codera, selon les mots, un seul domaine :

entrée	domaine	sous-domaine
<i>animateur-vedette</i>	<b>audiovisuel</b>	<b>radio ; télévision</b>
<i>apocalypse</i>	<b>religions</b>	<b>judaïsme ; christianisme</b>
<i>hystérectomie</i>	<b>médecine</b>	<b>chirurgie ; gynécologie</b>
<i>porte-cartons</i>	<b>art</b>	<b>dessin ; peinture artistique</b>

ou plusieurs :

entrée	domaine	sous-domaine
<i>artillerie de marine</i>	<b>militaire ; nautique</b>	<b>artillerie ; marine militaire</b>
<i>enfant à charge</i>	<b>sociologie ; finance</b>	<b>famille ; fiscalité</b>

On voit ainsi que le lien entre les deux rubriques n'a rien de systématique<sup>17</sup>. Il en résulte que les rattachements que nous proposons n'ont qu'une valeur indicative : ils représentent seulement les cas les plus fréquents. Il faut en réalité décider, pour chaque unité lexicale, du codage le plus pertinent. Domaines et sous-domaines ne sont pas redondants mais bien plutôt complémentaires : les deux rubriques apportent chacune des informations spécifiques et autorisent des extractions à des degrés différents de généralité.

---

<sup>17</sup> Le mot *enveloppe à fenêtre* en offre un bon exemple. Le codage des sous-domaines (**poste** et **correspondance**) ne pose guère de problème. Il est concevable par ailleurs, sur un plan général, de rattacher la poste au domaine des **services** et la correspondance à la **communication**. Mais la relation n'est pas transitive : quel sens y aurait-il à dire que *enveloppe à fenêtre* relève du domaine des "services" ?



## Bibliographie

- BUVET, Pierre-André (1996), « Levée d'ambiguïté et domaines d'emploi », *Bulag* n°21, Besançon, Université de Franche-Comté.
- DUBOIS, Jean & alii (1973), *Dictionnaire de linguistique*, Paris, Larousse.
- GALISSON, Robert & Daniel COSTE (1976), *Dictionnaire de didactique des langues*, Paris, Hachette.
- GIRARDIN, Chantal (1994), *Images de l'usage et attitudes sociales dans les dictionnaires classiques*, Thèse de doctorat, Université de Paris VIII.
- GLATIGNY, Michel (1994), « Les marques d'usage dans les dictionnaires français monolingues », *Les dictionnaires de langue. Méthodes et Contenus*, Actes du colloque « La journée des dictionnaires », Université de Cergy-Pontoise.
- GROSS, Gaston (1992), « Forme d'un dictionnaire électronique », *Actes du colloque : La station de traduction de l'an 2000*, Mons.
- GROSS, Gaston (1994), « Classes d'objets et description des verbes », *Langages* 115, Larousse, Paris.
- GROSS, Gaston (1995), « Une sémantique nouvelle pour la traduction automatique. Les classes d'objets », *La tribune des industries de la langue et de l'information électronique* n° 17-18-19, Paris.
- ISO~1807 (1990), *Terminologie. Vocabulaire, norme internationale*, Genève, ISO.
- KOCOUREK, Rostislav (1991), *La langue française de la technique et de la science*, Wiesbaden, Brandstetter Verlag.
- LE PESANT, Denis (1994), « Les compléments nominaux du verbe lire », *Langages* 115, Larousse, Paris.
- LE PESANT, Denis et Michel MATHIEU-COLAS (1998), « Introduction aux classes d'objets », *Langages* 131, Larousse, Paris.
- LERAT, Pierre (1995), *Les langues spécialisées*, Paris, Presses Universitaires de France
- MAROUZEAU, Jean (1969), *Lexique de la terminologie linguistique, français, allemand, anglais, italien*, Paris, Geuthner.
- MATHIEU-COLAS, Michel (1994), *Les mots français à trait d'union*, Paris, Didier Erudition.
- MATHIEU-COLAS, Michel (1998), « Illustration d'une classe d'objets : les voies de communication », *Langages* 131, Larousse, Paris.
- QUEMADA, Bernard & alii (1984), *Datations et documents lexicographiques, Matériaux pour l'histoire du vocabulaire français*, 2<sup>e</sup> série, vol. 2, CNRS, Paris, Klincksieck.
- REY-DEBOVE, Josette & Christine de BELLEFONDS. (1989), « Présentation », *Le Robert méthodique*, Paris, Dictionnaires Le Robert.

## ANNEXE

A titre d'illustration, nous présentons ici de brefs extraits de nos listes :

1. Domaines.
2. Sous-domaines élémentaires (d'autres sous-domaines peuvent être codés à l'aide d'un transdomaine ou d'un spécificateur : cf. *supra*, 2.1).
3. Principales correspondances entre domaines et sous-domaines.
4. Table de correspondance entre certaines rubriques lexicographiques et leur équivalent dans notre nomenclature.

### 1. LISTE DES DOMAINES CODÉS (extraits)

DOMAINE	CODE
administration	admin.
aéronautique	aéron.
agriculture	agric.
alimentation	alim.
.../	
commerce	comm.
communication (divers)	communic.
construction	constr.
croyances et idéologies	croy.-idéol.
.../	
droit et justice	droit-jus.
économie	écon.
écrit (divers)	écrit
édition et métiers du livre	livre
éducation	éduc.
.../	
transports (divers)	transp.
travaux et équipement ménagers	ménag.
vie quotidienne (divers)	vie quot.
voyages	voy.

## 2. LISTE DES SOUS-DOMAINES CODÉS (extraits)

SOUS-DOMAINES	CODE
accidents	accid.
acoustique	acoust.
acupuncture	acup.
aérodynamique	aérodyn.
aéronavale	aéronav.
aérospatiale	aérosp.
agronomie	agron.
alchimie	alchim.
algèbre	algèb.
allergologie	allerg.
.../	
caoutchouc	caoutch.
caractérologie	caractérol.
carbochimie	carbochim.
cardiologie	cardiol.
.../	
crustacés	crust.
cryptographie	cryptog.
cuirs	cuirs
cuisine (articles de)	cuis.
.../	
entomologie	entom.
entreprise	entrepr.
entretien et maintenance	entr.-maint.
environnement	environ.
épigraphie	épigr.
.../	
maçonnerie	maçon.
magnétisme	magnét.
mammalogie	mammal.
manutention	manut.
.../	
voirie	voirie
voitures hippomobiles	voit.hipp.
volcanologie	volcan.
water-polo	wat.pol.
zootechnie	zootechn.

### 3. PRINCIPALES CORRESPONDANCES ENTRE DOMAINES ET SOUS-DOMAINES (extraits)

SOUS-DOMAINE	DOMAINE
accidents	0
acoustique	physique (+ ...)
acupuncture	médecine
aérodynamique	physique (+ ...)
aéronavale	aéronautique + nautique + militaire
aérospatiale	aéronautique + espace
agronomie	agriculture
alchimie	croyances et idéologies + chimie
.../	.../
aménagement du territoire	économie + environnement et urbanisme
ameublement	ameublement et décoration
analyse mathématique	mathématiques
analyse numérique	mathématiques
anatomie	biologie
anatomie animale	zoologie
.../	.../

### 4. TABLE D'ÉQUIVALENCE (extraits)

Antiquité précolombienne	Antiquité (Amérique latine)
appareillage médical	médecine:EQUI / prothèses
arboriculture	arboriculture fruitière / sylviculture
architecture navale	construction navale
armée	militaire
art capillaire	coiffure / cosmétologie
art dentaire	odontologie / stomatologie
.../	.../
machines agricoles	agriculture:EQUI
magie	occultisme / prestidigitation
manège	équitation
.../	.../
sciences de l'espace	espace
sciences de la Terre	géosciences
sciences économiques	économie:THEO
.../	

On remarquera, dans ces derniers exemples, l'utilisation de deux transdomaines (cf. *supra*, 2.1.1.) : EQUI (équipement) et THEO (théorie).